Psychometric properties of the Chinese version of the SF-36 in older adults with diabetes in Beijing, China

By: Jie Hu, Kenneth J. Gruber, and Kuei-Hsiang Hsueh

Hu, J., Gruber, K. J. & Hsueh, K.H. (2010). Psychometric properties of the Chinese version of the SF-36 in older adults with diabetes in Beijing, China. *Diabetes Research and Clinical Practice*, 88 (3), 273-281. doi:10.1016/j.diabres.2010.03.005

Abstract:

Objective

This study investigated the psychometric properties of the 36-item Short-Form Health Survey (SF-36) (China version) in older Chinese with diabetes living in Beijing, China.

Methods

The SF-36 was administered to community-based sample of 182 older adults with diabetes living in Beijing. Data collection was conducted in face-to-face interviews. Reliability and validity were assessed using internal consistency, convergent and discriminant analyses. Exploratory principal components analyses (PCA) were conducted to compare the sample's response patterns with the hypothesized scale constructs.

Results

Item level validation of the scale supported the assumptions of the hypothesized structure. Internal consistency reliability (Cronbach's alpha >.70) of the subscales were acceptable except for the General Health subscale (.67). PCA confirmed general support of the two hypothesized dimensional factors and eight concepts (factors). The physical component summary (PCS) and the mental component summary (MCS) explained 62.26% of the variance and the eight factors components explained 67.39% of the variance. Known-group comparisons of scale scores indicated significantly higher levels of functionality for respondents with no blood pressure, heart, or depressive symptomatology problems.

Conclusions

The Chinese version of the SF-36 showed good reliability and validity and was culturally equivalent. The scale is appropriate for use with older Chinese adults with diabetes.

Keywords: Psychometrics; SF-36; Health-related quality of life; Older Chinese adults; Diabetes

Article:

1. Introduction

According to the International Diabetes Federation [1], diabetes has reached epidemic proportions in China. In 2007, 39.8 million of Chinese were diagnosed with type 2 diabetes. Outpacing most other countries, China has the second largest number of persons suffering from type 2 diabetes in the world, and this number is expected to increase to 59.3 million by 2025 [1]. Furthermore, it is estimated that over two-thirds of the population that has diabetes living in mainland China are currently undiagnosed [2].

Among the diagnosed population, the highest prevalence of diabetes is among older adults and people living in urbanized cities [1], [2], [3] and [4]. This is not surprising given the likelihood that older Chinese are more likely to experience health problems requiring medical attention and that there are more clinics and other sources of health providers to provide diagnoses in more populated regions.

Because of the high prevalence of diabetes, it is critical to identify approaches that might help to reduce the incidence of poorly managed diabetes and increase the likelihood that those with self care issues follow proscribed care management to maintain control of the disease. Because it is impractical for health care providers to track care management, alternative approaches are necessary. One such approach is to employ measures of health-related quality of life that can provide indications of impact of a disease on the individual's well-being [5]. Health-related quality of life has been defined as an individual's perception of physical and mental health [6]; it is a multidimensional construct that includes physical functioning, role limitations, mental health, social functioning and general health perceptions [7].

The measurement of health-related quality of life has been recognized as a reliable means for assessing the impact of chronic illness on individuals [8] and [9]. A number of studies have shown that low health-related quality of life scores are associated with mortality among patient population and individuals with type 2 diabetes [10] and [11]. Assessment of health-related quality of life in individuals with diabetes has been considered an important factor in predicting mortality [11]. Individuals with diabetes complications have reported poorer health-related quality of life scores than those without [12] and those with lower scores physical component health-related quality of life scores have a 2.1 times greater risk in increased mortality than persons with normal range scores [11].

As a health-related quality of life measure, the 36-item Short-Form Health Survey (SF-36) has an established reputation for measuring physical and mental health functioning/well-being in varied samples and settings [13], [14], [15], [16], [17] and [18]. The SF-36 comprises Likert-type and dichotomous items measuring two major components of health, each of which consists of four subscales. The physical component summary (PCS) includes physical functioning (PF), role physical (RP), bodily pain (BP), and general health perceptions (GH). The mental component summary (MCS) includes vitality (VT), social functioning (SF), role-emotion (RE), and mental health (MH).

The SF-36 has been shown to have very good psychometric properties, with high internal consistency and test-retest reliability values from studies of populations with a range of ages and health conditions in the U.S. [18] and [19] and worldwide [20], [21], [22] and [23]. But, only a few studies have investigated the properties of the SF-36 with Asian samples, primarily Chinese residing in mainland China. These studies have found the summary scales and subscales to have good construct, convergent, and discriminant validity. The findings have shown good correspondence of the items with the two summary scales (PCS and MCS), but less consistent with expected associations of individual items with their hypothesized subscales [24], [25], [26], [27] and [28]. Studies that have used the SF-36 for assessing health-related quality of life of Asian continent-based Chinese samples have, with few exceptions [28], focused on non-mainland Chinese respondents [24], [26] and [27]. Because of cultural and conditional relevance factors, it may be the case that scale will perform similarly among mainland older Chinese adults with diabetes. Therefore, it is important to investigate the use of the SF-36 with a sample from the Chinese mainland to determine if the instrument's psychometric properties perform in line with hypothesized factors and provide scale scores that are valid, culturally relevant and clinically useful.

The present study presents a report of the psychometric properties of the SF-36 with a sample of older Chinese with type 2 diabetes living in Beijing, China. Specifically, the study assessed the internal consistency, scale and subscale assumptions, factor analysis, item discriminant validity, and the clinical validation of scale and subscale scores (construct validity) based on correspondence with health status indicators of the respondents.

2. Subjects, materials and methods

2.1. Subjects, setting, and procedures

A convenience sample of 182 older adults with diabetes in Beijing, China was recruited for participation in this study over the course of the summers of 2005 and 2006. Inclusion criteria included being diagnosed with diabetes; aged 60 or older; oriented to place and time (without cognitive impairment); and able to speak Mandarin. Participants were recruited from 12 residential apartment complexes in two major districts of Beijing. Each residential apartment complex housed approximately 350–400 households. On the first floor of each complex was a residential center that was used by the complex's residents for social activities and health education. Flyers describing the study were posted on the community bulletin board of each complex. In addition, the administrators of each of the complexes informed (by phone or in person) all the residents they knew who had diabetes a week before the study team visited their complex. Data collection took 1–2 days at each complex. A total of 185 participants were recruited; three were excluded from the study due to a determination that they were not diagnosed with diabetes. Both verbal and written explanations of the study were provided to potential participants. After obtaining signed informed consent, the study questionnaires were administered along with collection of respondents' blood pressure, glucose level, weight, height, and waist circumference. The first author and native Chinese registered nurses in Beijing administered the surveys data in face-to-face interviews in a linguistically and culturally sensitive manner. Face-to-face interviews were chosen for this population because many older Chinese adults have low literacy levels and some of older adults were likely to have vision problems making completing a written survey difficult. The data collection session took about 40–50 min to complete. After completion of their participation, each participant received a set of towels as compensation for their time. Human subjects approval was obtained from the Institutional Review Board of the first author's University and from the local community centers in the neighborhoods in Beijing where respondents were recruited. The guarantee of the confidentiality was explained in oral and written form.

2.2. Measures

2.2.1. The 36-item Short-Form Health Survey (SF-36)

A Chinese version of the SF-36 (China) published by Quality Metric Health Outcomes Solutions [29] was used in the study. The SF-36 Health Survey was developed in the United States and has been translated, validated and established norms of the SF-36 in 14 countries, including Austria, Belgium, Canada, Denmark, France, Hong Kong, Italy, Japan, the Netherlands, Norway, Spain, Sweden, UK, and US (Chinese translation) in an International Quality of Life Assessment (IQLA) Project. Results generally supported the cross-cultural validity of the SF-36 Health Survey, making possible comparisons of health-related quality of life across countries and encouraging a wide use of this instrument [30]. Responses to scale items are made using a standard set of response choices with Likert scaling that are then combined to form eight summed subscales. Subscale scores are then transformed from normal scaling to a 0–100 standardized score, with higher scores indicating more positive health status, and a better health-related quality of life. Scoring is normed using published means, standard deviations, and aggregated weights. T-score transformations are used for the aggregated physical component summary (PCS) and mental component summary (MCS) scores to arrive at mean of 50 and a standard deviation of 10 [18] and [31]. Missing data are replaced by substituting individual mean scores for the non-missing items on the subscale to which the missing item response was associated [18], [19] and [32]. Data imputation was used for missing values replacement if respondents have less than one-half of the items of a subscale missing.

2.2.2. Geriatric Depression Scale-Short-Form

The 15-item Geriatric Depression Scale-Short-Form (GDS-S) [33] assesses depressive symptoms in older adults. Items are rated on a dichotomous scale scored 0 indicating no depression or 1 indicating depression. Items are summed with a total score between 0 and 4 indicating no or mild depression, 5–8 reflecting moderate depression, and 9 or greater representing severe depression. The sensitivity and specificity of the GDS-S have been reported to be satisfactory with elderly populations [34]. Reliability and convergent validity have been established among older adults [33] and minority populations [35] and [36]. Cronbach's alpha was .80 in this study. The GDS-S has been translated into Chinese and tested with Chinese populations with satisfactory reliability [37].

2.2.3. Demographic information and physiological measures

Demographic information was obtained using a questionnaire that asked for respondent's age, gender, education, income, marital status family history of diabetes, length of time since diagnosed with diabetes, self-reported diagnosed medical conditions, and use of medications, including traditional Chinese medicine. Respondent status of having a diagnosis of diabetes, heart disease, or hypertension was based on their response to the question: "Has your doctor ever told you that you have diabetes, or heart disease, or hypertension?" or if the respondent indicated he/she was currently taking medication for these health problems. Physiological measures including glucose, blood pressure, body mass index (BMI) and waist circumference were obtained through blood tests and body measurements.

2.3. Analyses

To test the construct and discriminant validity of the SF-36, several descriptive and correlational computations were performed [19], [32] and [38]. First, to determine the scale's ability to detect differences or change, the scoring range and percent of respondents achieving the highest (ceiling) and lowest (floor) possible scores for the summary dimensions and subscales were then calculated [19]. High percentages of ceiling or floor range values indicate low ability of a scale to detect changes over time.

Second, a correlation matrix of the individual items and scales was examined to determine how well each item related to its hypothesized subscale. To test this relationship, internal consistency reliability statistics (Cronbach's alphas) were computed. Evaluation criteria for internal consistency were based on item–item (.30–.70), item–subscale (.50–.65), item–total scale (≥.30), subscale–subscale (.40–.65), subscale–total scale (≥.55) correlation ranges. A minimum reliability coefficient of .70 has been recommended for group-level analyses [39].

Third, to assess convergent validity, correlations of each item with its hypothesized subscale corrected for overlap (all items except the target item) were computed to determine whether correlations exceeded the criterion of .40.

Fourth, item discriminant validity (the extent for which an item does not measure other constructs) was examined by whether the correlation of each item with its scale was greater (by a margin of two standard errors – approximately a 95% confidence interval) with its hypothesized scale than with the other scales making up the test. The standard error of a correlation coefficient is approximately equal to 1 divided by the square root of the sample size. The ratio of the percentage of successes in item discriminant validity tests to the total number of tests conducted is referred to as the scaling success rate [18].

Finally, the discriminant validity of each subscale was tested by comparing correlations of items from a subscale with its correlations to the other subscales. The correlation of items with their matched subscale should different significantly from correlations with the other subscales [18].

To test for construct validity and determine if the Chinese version of the SF-36 produced a factor pattern (subscales) similar to the hypothesized pattern of the standard English language version, an exploratory principal components analysis with varimax rotation of the individual items was conducted. A second order principal components analysis of the eight subscales to examine their correspondence with the two hypothesized summary dimensions PCS and MCS was conducted. Components associated with eigenvalues 1.0 or greater were selected. Item loadings of .50 and greater were used to represent the factors. Factor loadings ≥.40 were considered to be significant [39] and [40].

Known-group comparisons were then made for older adults with high, moderate, and low depressive symptoms measured by the Geriatric Depression Scale. Further comparisons were conducted with self-reports of heart disease and hypertension. Support for clinical validation of the scale is provided when known groups differ significantly on factors known to contribute to differences in health-related quality of life [41]. We expected that groups with moderate or high levels of depressive symptoms or self-reported heart disease or hypertension would have poorer health-related quality of life than groups with a low level of depressive symptoms and groups without heart disease or hypertension. One-way ANOVA and independent t-tests were used to test these hypotheses among groups [42].

3. Results

3.1. Characteristics of the sample

One hundred eighty-two older adults with a mean age of 69 (SD = 6.89) ranging from 60 to 90 years participated in the study. Most were female (64%) with less than a high school education (62%) and 28% had low incomes (<1000 RBM/month). The majority of participants reported having heart disease (53%) or hypertension (60%). On the GDS-S, most participants reported low levels of depressive symptoms (58%), but one-third (31%) were classified as having moderate depression, and a substantial minority (12%) were considered to be severely depressed.

3.2. Psychometric properties of the SF-36

3.2.1. Floor and ceiling effects

As shown in Table 1, the floor effects were modest for all the subscales, indicating that few respondents rated the items as applying extremely negative to themselves. Regarding ceiling effects however, both the SF and the RE subscales had substantial top score ratings; over a third of the respondents rated all the social functioning and role-emotional items as highly positive. For the RP and BP subscales, over one-fourth of the respondents rated items in the high positive range.

Table 1.
Floor and Ceiling of the SF-36 (N = 182).

Scale	Floor	Ceiling
Physical functioning (PF)	1.1	15.4
Role physical (RP)	4.9	25.8
Bodily pain (BP)	.5	26.4
General health (GH)	1.1	.5
Vitality (VT)	.0	.0
Social functioning (SF)	2.7	39.6
Role-emotional (RE)	2.7	36.9
Mental health (MH)	.0	.0

3.2.2. Internal consistency

All but two items met the item-to-hypothesized scale >.40 correlation on the GH subscale: GH1- self-rating of general health and GH1a – sick easier than others. Correlations between items and their hypothesized subscales were comparable across the subscales. Item-total correlations were positive, with coefficients ranging from .31 to .79 on the eight subscales. Both the PCS and MCS, and all subscales exceeded the recommended Cronbach's alpha level of .70 for internal consistency except for the GH which had a Cronbach's alpha of .67 (Table 2).

Table 2.
Item-scale correlation, internal consistency reliability and scale success rate (N = 182).

Scale	Range of item-total correlation	Reliability coefficient (%)	Success rate
Physical functioning (PF)	.49–.81	.92	98.8
Role physical (RP)	.75–.78	.89	100.0
Bodily pain (BP)	.73–.73	.85	100.0
General health (GH)	.31–.56	.67	100.0
Vitality (VT)	.45–.58	.74	100.0
Social functioning (SF)	.55–.55	.71	100.0
Role-emotional (RE)	.64–.79	.87	100.0

Scale	Range of item-total correlation	Reliability coefficient (%)	Success rate
Mental health (MH)	.42–.69	.78	100.0
Physical component summary (PCS)	.27–.73	.91	–
Mental component summary (MCS)	.47–.69	.89	–

3.2.3. Item discriminant validity

Item discriminant validity was evidenced by high significant correlations with their hypothesized subscales being two standard error scores higher on their hypothesized subscales than with other subscales across all 8 subscales with one exception. Item 3a – "vigorous activities" was correlated .59 with its hypothesized PF subscale scale score and .49 with the RP subscale score. The item discriminant success rate was 100% for seven subscales except for the PF, which had rate of 99% (Table 2) due to item 3a (vigorous activities).

3.2.4. Principal component analysis

The initial factor analysis yielded a 9-factor solution accounting for 70% of the variance. All but one factor had at least two item loadings of .50 or higher. Because we wanted to match the eight component solution of the standard English version of the scale, a second analysis was conducted restricting the results to an eight component solution. This analysis yielded a solution accounting for 67% of the variance. The results are presented in Table 3. The factor patterns generally corresponded with the eight hypothesized subscales. One notable exception was the alignment of General Health, Vitality and Mental Health items on two different components, suggesting some possible incongruity of the items making up these two hypothesized subscales for this sample.

Table 3.
Factor pattern of SF-36 (Chinese version) Items (principal components with varimax rotation).

	Components							
	F1	F2	F3	F4	F5	F6	F7	F8
Bodily pain								
BP1 (7)							.88	
BP2 (8)							.87	
General health								
GH1 (1)								.62
GH2 (11a)					.54	.51		
GH3 (11b)					.69			
GH4 (11c)					.59			
GH5 (11d)					.66			
Mental health								
MH1 (9b)						.63		
MH2 (9c)			.50	.40				
MH3(9d)			.73					
MH4 (9f)		.46	.41	.41				

	Components							
	F1	F2	F3	F4	F5	F6	F7	F8
MH5 (9h)			.65					
Physical functioning								
PF1 (3a)	.44			.41				
PF2 (3b)	.75							
PF3 (3c)	.83							
PF4 (3d)	.78							
PF5 (3e)	.65							
PF6 (3f)	.69							
PF7 (3g)	.80							
PF8 (3h)	.84							
PF9 (3i)	.78							
PF10 (3j)	.60							
Role-emotional								
RE1 (5a)		.82						
RE2 (5b)		.84						
RE3 (5c)		.68						
Role physical								
RP1 (4a)		.69						
RP2 (4b)	.41	.61						
RP3 (4c)	.46	.59						
RP4 (4d)	.46	.63						
Social functioning								
SF1 (6)							.60	
SF2 (10)						.62		
Vitality								
VT (9a)				.74				
VT (9b)				.69				
VT (9c)		.69						
VT (9d)		.81						

Loadings .50 and higher are bolded. BP = bodily pain, GH = general health, MH = mental health, PF = physical functioning, RE = role-emotional, RP = role physical, SF = social functioning, VT = vitality. F1–F8 = Factor 1–Factor 8

Determination of the subscales to adequately characterize the two major dimensions of health, the PCS and MCS, was tested. Principal components analysis limiting the solution to two factors produced a matrix accounting for 62% of the variance. As shown in Table 4, these results suggest that the Chinese version of the SF-36 has reasonable two-dimensionality. All subscales clearly loaded on their respective theoretical factors with the exception of the GH which unexpectedly loaded on the mental domain of health rather than physical domain of health. The RP cross-loaded on both the PCS and MCS, with slightly higher loadings on the MCS.

Table 4.
Factor pattern* of SF-36 subscales by hypothesized major dimensions.

Subscales	Summary scale dimensions	
	PCS	MCS
Physical functioning (PF)	.65	.42
Role physical (RP)	.57	.65
Bodily pain (BP)	.85	.00
General health (GH)	.25	.69
Vitality (VT)	.14	.79
Social functioning (SF)	.32	.60
Role-emotional (RE)	.36	.67
Mental health (MH)	.05	.86

PCS = physical component summary, MCS = mental component summary.
* Significant ($p < .05$) loadings appear in bold.

3.2.5. Known-group differences

Table 5 presents comparative results of SF-36 mean scale scores by heart problem, high blood pressure, and depression level status. Inspection of the table shows that with the exception of the BP scale, respondents with no heart problems and no high blood pressure reported significantly higher functioning scores on the SF-36 than respondents with heart problems or high blood pressure. Groups formed from scores on the GDS-S revealed that for the eight subscales and two-dimensional scales, with the exception of the BP scale, respondents who reported low or no depressive symptomatology had higher functioning scores on the SF-36. For all the subscales, the low or no depression group had significantly higher scores than the moderate to severe depressive symptomatology group.

Table 5.
Comparison of SF-36 mean scale scores by heart problem, high blood pressure, and depression level status.

	PF*	RP	BP	GH	VT	SF	RE	MH	PCS	MCS
Heart problems										
No (N = 86)	82.1c (19.7)d	75.6 (26.2)	68.5 (25.5)	53.1 (21.8)	64.0 (17.6)	83.1 (22.1)	78.6 (24.0)	70.7 (14.2)	69.8 (17.4)	74.1 (15.3)
Yes (N = 96)	65.2 (29.2)	56.0 (32.6)	61.2 (27.4)	46.0 (21.1)	56.2 (17.4)	72.8 (27.4)	68.9 (29.1)	64.2 (16.0)	57.1 (19.8)	65.5 (17.5)
High blood pressure										
No (N = 75)	84.6 (16.8)	78.5 (25.3)	67.0 (24.9)	56.9 (21.5)	65.1 (14.9)	82.7 (22.9)	81.6 (23.7)	71.5 (13.5)	71.7 (16.6)	75.2 (15.6)
Yes (N = 107)	65.1 (29.0)	56.0 (31.8)	63.0 (27.8)	44.1 (20.3)	56.2 (18.9)	74.2 (26.8)	67.8 (28.1)	64.3 (16.2)	57.1 (19.5)	65.6 (16.9)

	PF*	RP	BP	GH	VT	SF	RE	MH	PCS	MCS
Depression level**										
None (N = 104)	78.5a (23.0)	75.5a (25.4)	67.4 (25.3)	54.9ab (20.4)	65.7a (14.8)	82.6a (19.7)	80.7a (21.4)	72.2a (11.5)	69.1a (16.7)	75.3a (12.3)
Mild (N = 56)	69.5a (28.3)	56.1a (33.1)	63.5 (27.5)	43.6a (20.4)	55.9a (18.2)	75.9b (27.7)	66.7a (30.7)	64.6a (15.1)	56.2a (20.0)	65.8a (17.7)
Moderate or Severe (N = 21)	56.8a (30.8)	40.2a (31.7)	52.3 (28.1)	38.1b (23.7)	43.6a (15.9)	60.1ab (35.3)	56.0a (31.9)	50.1a (20.3)	46.9a (21.1)	52.4a (20.0)

Means bolded are significantly different; Means with same superscripts are significantly different ($p < .05$).
c Mean
d Standard Deviation
* PF = physical functioning, RP = role physical, BP = bodily pain, GH = general health, VT = vitality, SF = social functioning, RE = role-emotional, MH = mental health, PCS = physical component summary, MCS = mental component summary.
** Depression levels based on responses to the short-form of the Geriatric Depression Scale (GDS-S).

4. Discussion

This study examined the psychometric properties of the version of the Chinese SF-36 [29] among older Chinese adults with type 2 diabetes. The findings of this study support the validity and reliability of the scale as a culturally adequate health-related quality of life measure for older adults residing in mainland China. Data completeness, data quality and the psychometric properties of the scale were similar to those reported in other studies conducted with Chinese samples from or around China [24], [27], [28] and [43], suggesting that the Chinese SF-36 can be used in cross-cultural studies of older Chinese adults with diabetes.

Responses to the Chinese SF-36 version 2 had lower percentage of missing data in previous studies with Chinese populations [25] and [27]. This may be partly due to this version of the scale but more likely due to the fact that unlike the other studies, we used face-to-face interviews to collect scale data.

Floor effects representing percentage of respondents reporting lowest possible scores on the subscales were very low (less than 5% on any one subscale). Ceiling effects representing the highest possible score on items suggest that several of the SF and RE items may have low discriminative power; over a third of the sample rated these items with the highest possible positive ratings. However, our findings are consistent with floor and ceiling effects reported in previous studies of Chinese and U.S. populations [24], [27], [28], [38] and [44], so this may be more of an artifact of the dimension being measured than an issue of discriminative ability. The current study found that the highest ceiling effect on the SF subscales, which is consistent with findings reported in a Chinese population in Hong Kong [44]. The high percentage of ceiling effects on the SF and RE may be a reflection of the characteristics of the sample, i.e., the majority reported relatively low depressive symptoms and so their perceived social functioning might have been quite high. Also, in Chinese culture, emotional problems are stigmatized and it is less likely that individuals will admit to relationship problems (measured in both SF and RE), resulting in high ratings indicating “no problem” on these items. This is reflected in traditional Chinese cultural concepts and beliefs about fatalism (fate or destiny) that encourage Chinese people to withstand hardship, be tolerant of distressing circumstances, and accept depression [45] and [46]. Mental illness is stigmatized in traditional Chinese culture as loss of face or causing shame for the family members so admissions of emotional problems are often not reported [47]. Chinese culture emphasizes the importance of self-control of the individual's feelings and desires [48], which also might influenced their high scores on the mental health components.

The results of the scaling analyses supported the hypothesized scale structure of the SF-36 in Chinese older adults with diabetes. Scaling success rates are high for all the subscales with the PF1 (item 3a) having the lowest scaling success rate (98.8%). The low scaling success rate of the PF1 was probably due to the culturally unsuitable question that asked about vigorous activities, for which the majority of older Chinese adults don't typically engage. Vigorous activities, such as running and participating strenuous sports are not traditional activities for elders. A recent epidemiological study with randomized sampling reported that only 22% of urban Chinese adults aged 35–74 in mainland China were physically active and participation of moderate or vigorous leisure time physical activity was even lower with only 7.8% [49]. In traditional Chinese society, slow-pace, moderate exercises such as brisk walking, Chi and Qi Gong are the more usual physical activities of older adults. The association of the PF1 with the RF suggests that participation in vigorous activities might be perceived conceptually as limitations in role physical among older Chinese adults with diabetes, for example limitations in types of work or activities and difficulties in performance of work or activities. Future studies will need to consider whether and how to modify this item for use with older Chinese populations.

Internal consistency of the two summary scales, PCS and MCS, were highly consistent with findings reported for both Chinese and U.S. populations [32] and [43]. Psychometric testing results for the eight subscales indicated generally good internal consistency reliability except for the general health and social functioning subscales. The low internal consistency reliability on general health noted in this population might be explained by the fact that in Chinese culture health is always viewed as regaining health and wellness, counterbalancing illness, searching balance and maintain equilibrium between yin and yang [50] and [51]. Having diabetes can be considered an imbalance or disequilibrium under traditional Chinese yin and yang principle. Older Chinese adults with diabetes may see their health as poorer compared with health of people without chronic conditions. Moreover, Chinese people are not inclined to say that their health is excellent because a widely believed superstition is that boasting about one's health may result in bad things happening [44]. In addition, there is a reluctance in older Chinese adults to conceptualize one's health compared with the health of others (GH2 [to get sick a little easier than other people] and GH3 [as healthy as anybody I know]). This suggests that the specific content of these questions or the way they are being asked may not be culturally appropriate for older Chinese adults. Low internal consistency of the SF scale also has been reported in other studies of Chinese populations [24], [27] and [28]. Cultural differences in the way the Chinese and Americans value social-related functioning on social and family activities might also account for lower reliability of scores for the Chinese sample on this scale. The impact of cultural difference in health perception and expression of mental health suggests that future cross-cultural studies are needed to examine more closely the relevance and appropriateness of the mental health components of the SF-36.

The principal components analysis of individual items and the eight subscales produced reasonably good matching of items to their hypothesized subscales and subscales to the hypothesized two summary dimensions. One exception from the individual item principal component analysis results was that the hypothesized general health, vitality and mental health items were associated with two distinct components rather than a single dimension. This finding is consistent, however, with two other studies reported principle component analysis of the SF-36 with Asian populations [26] and [28]. Our findings showed that VT3 (feel worn out) and VT4 (feel tired) loaded as the same factor components with MH4 (down hearted and depressed) and MH5 (been happy [reverse coded]) suggesting that the construct of the VT and MH subscales are related in Chinese adults with diabetes. Similar findings also have been reported among Taiwanese and Chinese Americans [27] and [52]. In the view of traditional Chinese medicine, depression related fatigue is often due to a lack of vital energy [52]. The loading of the GH subscale on mental domain of health was consistent with reports involving Chinese, Malays and Indians samples in Singapore [26]. Because of the differences in Western versus Asian cultural context surrounding health, it should not be surprising that the items on the GH subscale (general health status, sick easier, healthy as others, expect health to get worse, health is excellent) might be perceived differently. Additional research is needed to determine the importance of these cultural differences as they relate to health-related quality of life and the impact of chronic diseases like diabetes.

The Chinese SF-36 demonstrated good construct validity using known-group comparisons on both the eight subscales and the PCS and MCS scores. Significant differences were found between groups with low and moderate as compared with higher depressive symptoms. Also, older adults with self-reported heart disease and hypertension had significantly poorer health-related quality of life in all eight domains of health and on the two summary scores. Our findings indicate that the Chinese SF-36 adequately discriminates between groups of individuals known to differ by types of medical conditions. Similarly, Lam et al. [43] found that Chinese individuals with heart disease, hypertension, and diabetes had significantly lower PCS scores than individuals without these chronic conditions. But, they also found no significance differences associated with the MCS.

There are several notable limitations to this study. One, it involved non-probability sampling which might limit the representativeness of the population of older Chinese with diabetes and thus the generalizability of the findings to the current population of older Chinese with diabetes. Two, SF-36 scores were calculated based on general U.S. population norms [7], which may not be representative of the Chinese population. There are no published norms available for the Chinese population. Because our purpose was to explore the conceptual validity of the scale using a mainland Chinese sample, comparison to the U.S. norms provided the appropriate reference frame. Other studies have also used U.S. norms to demonstrate functional equivalence of a Chinese (mainland) version of the SF-36 for the purposes of establishing cross-cultural validity [25]. Three, test and retest reliability also was not assessed in the study for establishment of stability of the SF-36. However, test-retest reliability has been established in Chinese patients with myocardial infarction [25] and [28]. Also, in contrast to most other studies using the SF-36, we administered the instrument in face-to-face interviews which could have affected the respondents' health assessments. However, the general agreement of our findings with other studies suggest that any social desirability effect bias influencing respondents to give responses that are socially acceptable in Chinese culture was likely minimal.

In summary, a psychometric analysis of the Chinese version of the SF-36 was conducted and the results show evidence to support the hypothesized 2-factor structure comprising eight subscales in our study. The results demonstrated good feasibility, acceptability, reliability, and validity. In addition, the Chinese version of the SF-36 performed well in discriminating participants with high versus no or low blood pressure, those with and without, heart problems, and those with no and low and moderate to severe levels of depressive symptomatology. Overall, the findings suggest that the Chinese version of the SF-36 is conceptually equivalent to the English version of the SF-36, indicating that it is a culturally adequate tool for measuring health-related quality of life in older Chinese with diabetes in mainland China.

Cultural differences in the general health subscale and differences in conceptualization of the vitality and mental health subscale need to be further tested in the Chinese cultural context with a larger and diverse Chinese population. Further testing of validity, culturally equivalence, sensitivity to change, and clinical significant differences of the Chinese SF-36 is needed.

Conflict of interest

No conflict of interest has been declared by the authors.

Acknowledgements

Support for this project was provided by Gamma Zeta Chapter of Sigma Theta Tau International Honor Society of Nursing, UNCG, School of Nursing and UNCG Kohler travel fund.

References

[1] International Diabetes Federation, Diabetes Atlas. (2009), available from http://www.eatlas.idf.org/index4352.html.
[2] K.C. Wong and Z. Wang, Prevalence of type 2 diabetes mellitus of Chinese populations in mainland China, Hong Kong, and Taiwan, Diabetes Res. Clin. Pract. 73 (2006), pp. 126–134.
[3] J.C. Chan, V. Malik, W. Jia, T. Kadowaki, C.S. Yajnik and K.H. Yoon et al., Diabetes in Asia: epidemiology, risk factors, and pathophysiology, JAMA 301 (2009), pp. 2129–2140.
[4] S. Wild, G. Roglic, A. Green, R. Sicree and H. King, Global prevalence of diabetes: estimates for the year

2000 and projections for 2030, Diabetes Care 27 (2004), pp. 1047–1053.
[5] A.R. Mailler, A.A. Kaptein, J.C.M. de Haes and W. Everaerd, Assessing quality of life in chronic non-specific lung disease – A review of empirical studies published between 1980 and 1994, Qual. Life Res. 5 (1996), pp. 287–301.
[6] Centers for Disease Control & Prevention, Health-related quality of life. (2006), available from http://www.cdc.gov/hrqol/index.htm.
[7] J.E. Ware, M. Kosinski and S.K. Keller, SF-36 Physical and Mental Health Summary Scales: A User's Manual, Health Institute New England Medical Center, Boston, MA (1994).
[8] K.T. Ashing-Giwa, The contextual model of HRQoL: a paradigm for expanding the HRQOL framework, Qual. Life Res. 14 (2005), pp. 297–307.
[9] G.H. Guyatt, D.H. Feeny and D.L. Patrick, Measuring health-related quality of life, Ann. Intern. Med. 118 (1993), pp. 622–629.
[10] V.S. Fan, D.H. Au, M.B. McDonell and S.D. Filn, Intraindividual change in SF-36 in ambulatory clinic primary care patients predicted mortality and hospitalizations, J. Clin. Epidemiol. 57 (2004), pp. 277–283.
[11] N. Kleefstra, G. Landman, S.T. Houweling, L.J. Ubink-Veltmaat, S.J. Lotenberg and B. meyboom-deJong et al., Prediction of mortality in type 2 diabetes from health-related quality of life (ZODIAC-4), Diabetes Care 31 (2008), pp. 932–933.
[12] J.T. Coffey, M. Brandle, H. Zhou, D. Marriott, R. Burke and B.P. Tabael et al., Valuing health-related quality of life in diabetes, Diabetes Care 25 (2002), pp. 2238–2243.
[13] S.C. Haffer, S.E. Bowen, E.D. Shannon and B.M. Fowler, Assessing beneficiary health outcomes and disease management initiatives in Medicare, Disease Manage. Health Outcome 11 (2003), pp. 111–124. (8)
[14] A.S. Russell, B. Conner-Spady, A. Mintz and W.P. Maksymowych, The responsiveness of generic health status measures as assessed in patients with rheumatoid arthritis receiving infliximab, J. Rheumatol. 30 (2003), pp. 941–947.
[15] A. Steptoe, A. Mohabir, N.G. Mahon and W.J. McKenna, Health-related quality of life and psychological well-being in patients with dilated cardiomyopathy, Heart 83 (2000), pp. 645–650.
[16] V. Strand, P. Tugwell, C. Bombardier, A. Maetzel, B. Crawford and C. Dorrier et al., Function and health-related quality of life: results from a randomized controlled trial of leflunomide versus methotrexate or placebo in patients with active rheumatoid arthritis. Leflunomide Rheumatoid Arthritis Investigators Group, Arthritis Rheum. 42 (1999), pp. 1870–1878.
[17] D.R. Walker, D.L. Landis, P.M. Stern and R.P. Vance, Disease management positively affects patient quality of life, Managed Care Interface 16 (2003), pp. 56–60.
[18] J.E. Ware, M. Kosinski, J.B. Bjorner, D.M. Turner-Bowker, B. Gandek and M.E. Maruish, User's Manual for the SF-36v2 Health (second ed.), Quality Metric Incorporated, Lincoln, RI (2007).
[19] C.A. McHorney, J.E. Ware, J.F. Lu and C.D. Sherbourne, The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups, Med. Care 32 (1994), pp. 40–66.
[20] C. Chittleborough, K. Baldock, A. Taylor and P. Phillips, The North West Adelaide health study T: Health Status assessed by the SF-36 along the diabetes continuum in an Australian population, Qual. Life Res. 15 (2006), pp. 687–694.
[21] I. Ohsawa, T. Ishida, Y. Oshida, K. Yamanouchi and Y. Sato, Subjective health values of individuals with diabetes in Japan: comparison of utility values with the SF-36 scores, Diabetes Res. Clin. Pract. 62 (2003), pp. 9–16. Article
[22] W. Schlotz, P. Ambery, H.E. Syddall, S.R. Crozier, A.A. Sayer and C. Cooper et al., Specific associations of insulin resistance with impaired health-related quality of life in the Hertfordshire Cohort Study, Qual. Life Res. 16 (2007), pp. 429–436.
[23] R.J. Tapp, D.W. Dunstan, P. Phillips, A. Tonkin, P.Z. Zimmet and J.E. Shaw, Association between impaired glucose metabolism and quality of life: Results from the Australian diabetes obesity and lifestyle study, Diabetes Res. Clin. Prac. 74 (2006), pp. 154–161.
[24] J.L. Fuh, S.J. Wang, S.R. Lu, K.D. Juang and S.J. Lee, Psychometric evaluation of a Chinese (Taiwanese) version of the SF-36 health survey amongst middle-aged women from a rural community, Qual. Life Res. 9 (2000), pp. 675–683.

[25] L. Li, H.M. Wang and Y. Shen, Chinese SF-36 Health Survey: translation, cultural adaptation, validation, and normalization, J. Epidemiol. Community Health 57 (2003), pp. 259–263.
[26] J. Thumboo, K.Y. Fong, D. Machin, S.P. Chan, K.H. Leong and P.H. Feng et al., A community-based study of scaling assumptions and construct validity of the English (UK) and Chinese (HK) SF-36 in Singapore, Qual. Life Res. 10 (2001), pp. 175–188.
[27] H.M. Tseng, J.F. Lu and B. Gandek, Cultural issues in using the SF-36 health survey in Asia: results from Taiwan, Health Qual. Life Outcomes 1 (7) (2003), pp. 1–9.
[28] W. Wang, V. Lopez, C. Ying and D. Thompson, The psychometric properties of the Chinese version of the SF-36 health survey in patients with myocardial infarction in mainland China, Qual. Life Res. 15 (2006), pp. 1525–1531.
[29] Quality Metric Health Outcomes Solutions. (2005), available from http://www.qualitymetric.com/products/license.
[30] B. Gandek and J.E. Ware Jr., Methods for validating and norming translations of health status questionnaires: the IQOLA Project approach. International quality of life assessment, J. Clin. Epidemiol. 51 (1998), pp. 953–959.
[31] J.E. Ware, K.K. Snow, M. Kosinski and B. Gandek, SF-36 Health Survey Manual and Interpretation guide., The Health Institute, New England Medical Center, Boston, MA (1993).
[32] C.A. McHorney, J.E. Ware and A.E. Raczek, The MOS 36-item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs, Med. Care 31 (1993), pp. 247–263. (2430)
[33] J.L. Sheikh and J.A. Yesavage, Geriatric depression scale (GDS) recent evidence and development of a shorter version, Clin. Gerontologist 5 (1986), pp. 165–173.
[34] P.G. Stiles and J.F. McGarrahan, The geriatric depression scale: a comprehensive review, J. Clin. Geropsychol. 4 (1998), pp. 89–110.
[35] J.M. Flacker and L. Spiro, Does question comprehension limit the utility of the geriatric depression scale in older African Americans?, J. Am. Geriatr. Soc. 51 (2003), pp. 1511–1512.
[36] Y. Jang, A. Borenstein-Graves, W.E. Haley, B.J. Small and J.A. Mortimer, Determinants of a sense of mastery in African American and white older adults, J. Gerontol. B Psychol. Sci. Soc. Sci. 58 (2003), pp. S221–224.
[37] K.W. Boey and H.F. Chiu, Assessing psychological well-being of the old-old: a comparative study of GDS-15 and GHQ-12, Clin. Gerontologist 19 (1998), pp. 65–75.
[38] B. Gandek, S.J. Sinclair, M. Kosinski and J.E. Ware Jr., Psychometric evaluation of the SF-36Â® Health Survey in medicare managed care, Health Care Finance Rev. 25 (2004), pp. 5–25.
[39] M.W. Nunnally and I.H. Bernstein, Psychometric Theory (3rd ed.), McGraw-Hill, New York (1994).
[40] B. Tabachnik and L. Fidell, Using Multivariate Statistics (4th ed.), Allyn and Bacon, Boston, MA (2001).
[41] J.E. Ware and B. Gandek, Methods for testing data quality, scaling assumptions, and reliability: the IQOLA project approach, J. Clin. Epidemiol 51 (1998), pp. 945–952.
[42] SPSS for Windows. Rel. 15.0 ed. SPSS Inc, Chicago, 2006.
[43] C.L. Lam, E.Y. Tse, B. Gandek and D.Y. Fong, The SF-36 summary scales were valid, reliable, and equivalent in a Chinese population, J, Clin. Epidemiol. 58 (2005), pp. 815–822.
[44] C.L. Lam, B. Gandek, X.S. Ren and M.S. Chan, Tests of scaling assumptions and construct validity of the Chinese (HK) version of the SF-36 Health Survey, J. Clin. Epidemiol. 51 (1998), pp. 1139–1147. Article
[45] G. Parker, G. Gladstone and K.T. Chee, Depression in the planet's largest ethnic group: the Chinese, Am. J. Psychiatry 158 (2001), pp. 857–864.
[46] J.M. Xu, Some issues in the diagnosis of depression in China, Can. J. Psychiatry 32 (1987), pp. 368–370.
[47] T. Ju-K'ang, Traditional Chinese beliefs and attitudes towards mental illness. In: Tseng, Editor, Chinese Cultural And Mental Health, Academic Press, Orlando (1985), pp. 67–84.
[48] M. Asadi-Lari, M. Tamburini and D. Gray, Patients' needs, satisfaction, and health-related quality of life: towards a comprehensive model, Health Qual. Life Outcome 2 (2004), p. 32 10.1186/1477-7525r-r2-3.
[49] P. Muntner, D. Gu, R.P. Wildman, J. Chen, W. Qan and P.K. Whelton et al., Prevalence of physical activity among Chinese adults: results from the international collaborative study of cardiovascular disease in Asia, Am. J. Public Health 95 (2005), pp. 1631–1636.

[50] Y. Chen, Chinese values, health and nursing, J. Adv. Nurs. 36 (2001), pp. 270–273.
[51] D.W.L. Lai and S. Surood, Chinese health beliefs of older Chinese in Canada, J. Aging Health 21 (2009), pp. 38–62.
[52] D.F. Chang, C. Chun, D.T. Takeuckhi and H. Shen, SF-36 Health Survey: tests of data quality, scaling assumptions, and reliability in a community sample of Chinese Americans, Med. Care 38 (2000), pp. 542–548